

**CLAIMS**

1. An intonation generation method for generating an intonation of text of speech synthesized by a computer having  
5 a memory location associated therewith, the method comprising:
  - estimating an outline of an intonation of the synthesized speech based on language information of the text and storing an estimation result in the memory;
  - reading out the estimation result of the intonation from  
10 the memory; and
  - selecting an intonation pattern from a database storing intonation patterns of actual speech based on the outline of the intonation, and defining the selected intonation pattern as the intonation pattern of the text.
- 15 2. The intonation generation method according to claim 1, wherein the outline of the intonation is estimated based on prosodic categories classified by the language information of the text.
- 20 3. The intonation creation method according to claim 1, wherein a frequency level of the selected intonation pattern is adjusted based on the estimated outline of the intonation after selecting the intonation pattern.
- 25 4. An intonation generation method for generating an intonation of text in a speech synthesized by a computer having an associated memory, the method comprising the steps of:

for each assumed accent phrase of the text being synthesized,

estimating an outline of the intonation for each assumed accent phrase and storing an estimation result in the memory;

5 reading out the estimated outline of the intonation for each assumed phrase, selecting intonation patterns from a database accumulating intonation patterns of an actual speech based on the outline of the intonation, and storing a selection result in the memory; and

10 reading out the selected intonation pattern for each assumed accent phrase from the memory, and connecting the intonation pattern to another.

5. The intonation generation method according to claim 4,  
15 wherein, in a case of estimating an outline of an intonation of a predetermined assumed accent phrase, when another assumed accent phrase is present immediately before the predetermined assumed accent phrase in the text, the step of estimating an outline of the intonation and storing an estimation result in a memory estimates the outline of the intonation of the predetermined assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase immediately therebefore.

25 6. The intonation generation method according to claim 4,  
wherein, when the assumed accent phrase is present in a phrase of a speech recorded in advance, the phrase being stored in a predetermined storage device, the step of estimating an outline

of the intonation and storing an estimation result in a memory acquires information concerning an intonation of a portion corresponding to the assumed accent phrase of the phrase from the storage device, and stores the acquired information as an 5 estimation result of an outline of the intonation in the memory.

7. The intonation generation method according to claim 6, wherein the step of estimating an outline of the intonation and storing an estimation result in a memory includes the steps of:

10 when another assumed accent phrase is present immediately before a predetermined assumed accent phrase in the text, estimating an outline of an intonation of the assumed accent phrase based on the estimation result of an outline of an intonation for the other assumed accent phrase immediately 15 therebefore; and

when another assumed accent phrase corresponding to the phrase of the speech recorded in advance, the phrase being stored in the predetermined storage device, is present immediately after the predetermined assumed accent phrase in the text, estimating 20 the outline of the intonation of the assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase immediately thereafter.

8. The intonation generation method according to claim 6, 25 wherein, when another assumed accent phrase corresponding to the phrase of the speech recorded in advance, the phrase being stored in the predetermined storage device, is present either before or after a predetermined assumed accent phrase in the text, the

step of estimating an outline of the intonation and storing an estimation result in a memory estimates an outline of an intonation for the assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase

5 corresponding to the phrase of the recorded speech.

9. The intonation generation method according to claim 4, wherein the step of selecting an intonation pattern and storing a selection result in the memory includes the steps of:

- 10 from among intonation patterns of an actual speech accumulated in the database, selecting an intonation pattern in which the outline is close to the outline of the intonation of the assumed accent phrase based on the distance from the starting point to the termination point; and
- 15 from among the selected intonation patterns, selecting an intonation pattern in which the distance of a phoneme class for the assumed accent phrase is smallest.

10. A speech synthesis apparatus for performing a

20 text-to-speech synthesis, comprising:

- a text analysis unit for analyzing text as a processing target and acquiring language information therefrom;
- a database for storing intonation patterns of actual speech;
- 25 a prosody control unit for generating a prosody for audibly outputting the text; and
- a speech generation unit for generating speech based on the prosody generated by the prosody control unit,

wherein the prosody control unit includes:

an outline estimation section for estimating an outline of an intonation for each assumed accent phrase configuring the text based on language information acquired by the text analysis

5 unit;

a shape element selection section for selecting an intonation pattern from the database based on the outline of the intonation, the outline having been estimated by the outline estimation section; and

10 a shape element connection section for connecting the intonation pattern for each assumed accent phrase to the intonation pattern for another assumed accent phrase, each intonation pattern having been selected by the shape element selection section, to generate an intonation pattern of an entire  
15 body of the text.

11. The speech synthesis apparatus according to claim 10, wherein the outline estimation section defines the outline of the intonation at least by a maximum value of a frequency level  
20 in a segment of the assumed accent phrase and relative level offsets in a starting end and termination end of the segment.

12. The speech synthesis apparatus according to claim 10, wherein the shape element selection section selects an  
25 intonation pattern approximate in shape to the outline of the information, the outline having been estimated by the outline intonation section, among the intonation patterns of the actual speech, the intonation patterns having been accumulated in the

database.

13. The speech synthesis apparatus according to claim 10,  
wherein the shape element connection section connects the  
5 intonation pattern for each assumed accent phrase to the other,  
the intonation pattern having been selected by the shape  
element selection section, after adjusting a frequency level of  
the assumed accent phrase based on the outline of the intonation,  
the outline having been estimated by the outline estimation  
10 section.

14. The speech synthesis apparatus according to claim 10,  
further comprising another database which stores information  
concerning intonations of a speech recorded in advance,  
15 wherein, when the assumed accent phrase is present in a  
recorded phrase registered in the other database, the outline  
estimation section acquires information concerning an  
intonation of a portion corresponding to the assumed accent  
phrase of the recorded phrase from the other database.

20 15. A speech synthesis apparatus for performing a  
text-to-speech synthesis, comprising:  
· a text analysis unit which analyzes text which is an object  
of processing and acquires language information therefrom;  
25 a plurality of databases prepared based on speech  
characteristics, the databases accumulating a plurality of  
intonation patterns of actual speech;  
· a prosody control unit which generates a prosody for

audibly outputting the text by use of the intonation patterns accumulated in the database; and

a speech generation unit which generates a speech based on the prosody generated by the prosody control unit,

5 wherein a speech synthesis on which the speech characteristics are reflected is performed by use of the databases in a switching manner.

16. A speech synthesis apparatus for performing a  
10 text-to-speech synthesis, comprising:

a text analysis unit which analyzes text as which is the object of processing, and acquires language information therefrom;

15 a first database which stores information concerning speech characteristics;

a second database which stores information concerning a waveform of a speech recorded in advance;

a synthesis unit selection unit which selects a waveform element for a synthesis unit of the text; and

20 a speech generation unit which generates a synthesized speech by coupling the waveform element selected by the synthesis unit selection unit to the other,

wherein the synthesis unit selection unit selects the waveform element for the synthesis unit of the text, the synthesis  
25 unit corresponding to a boundary portion of the recorded speech, from the first and second database.

17. A voice server for providing a content of a speech dialogue type in response to an access request made through a telephone network, comprising:

5 a speech synthesis engine for synthesizing a speech to be outputted to the telephone network; and

a speech synthesis engine for recognizing a speech received through the telephone network,

10 wherein the speech synthesis engine for recognizing a speech estimates an outline of an intonation for each assumed accent phrase configuring text based on language information of the text, the language information being obtained by executing an application, selects an intonation pattern from a database accumulating information patterns of an actual speech based on the estimated outline of the intonation for each assumed accent 15 phrase, connects the selected intonation pattern for each assumed accent phrase to another to generate an intonation pattern for the text, and synthesizes the speech based on the intonation pattern to output the synthesized speech to the telephone network.

20

18. A program for controlling a computer to generate an intonation in a speech synthesis, the program allowing the computer to execute:

25 processing of receiving language information of text as a target of the speech synthesis, estimating an outline of an intonation for each assumed accent phrase configuring the text based on the language information, and storing an estimation result in a memory;

processing of reading out the estimated outline of the intonation for each assumed accent phrase from the memory, selecting an intonation pattern from a database accumulating intonation patterns of an actual speech based on the outline of  
5 the intonation, and storing a selection result in the memory; and

processing of reading out the selected intonation pattern for each assumed accent phrase from the memory to connect the read out intonation pattern to the other, and outputting the  
10 connected intonation patterns as an intonation pattern for the text.

19. The program according to claim 18, wherein the processing of estimating an outline of an intonation and storing  
15 an estimation result in the memory, the processing being allowed by the program to be executed, includes processing of, in a case of estimating an outline of an intonation of a predetermined assumed accent phrase, when another assumed accent phrase is present immediately before the assumed accent phrase in the  
20 text, estimating the outline of the intonation of the predetermined assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase immediately therebefore.

25 20. The program according to claim 18, wherein, when the assumed accent phrase is present in a phrase of a speech recorded in advance, the phrase being stored in a predetermined storage device, the processing of estimating an outline of an intonation

and storing an estimation result in a memory, the processing being allowed by the program to be executed, acquires information concerning an intonation of a portion corresponding to the assumed accent phrase of the phrase from the storage device, and stores 5 the acquired information as an estimation result of an outline of the intonation in the memory.

21. The program according to claim 20,  
wherein the processing of estimating an outline of an 10 intonation and storing an estimation result in a memory, the processing being allowed by the program to be executed, includes:

processing of, when another assumed accent phrase is present immediately before a predetermined assumed accent phrase 15 in the text, estimating an outline of an intonation of the assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase; and

processing of, when another assumed accent phrase corresponding to the phrase of the speech recorded in advance, 20 the phrase being stored in the predetermined storage device, is present immediately after the predetermined assumed accent phrase in the text, estimating the outline of the intonation of the assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase 25 immediately thereafter.

22. The program according to claim 20, wherein, when another assumed accent phrase corresponding to the phrase of the

speech recorded in advance, the phrase being stored in the predetermined storage device, is present at least one of before and after a predetermined assumed accent phrase in the text, the processing of estimating an outline of an intonation and  
5 storing an estimation result in a memory, the processing being allowed by the program to be executed, estimates an outline of an intonation for the assumed accent phrase based on an estimation result of an outline of an intonation for the other assumed accent phrase corresponding to the phrase of the recorded  
10 speech.

23. The program according to claim 18, wherein the processing of selecting an intonation pattern, the processing being allowed by the program to be executed, selects an  
15 intonation pattern approximate in shape to the estimated outline of the information among the intonation patterns of the actual speech, the intonation patterns having been accumulated in the database.

20 24. A program for controlling a computer to perform a text-to-speech synthesis, the program allowing the computer to function as:

text analysis means for analyzing text as a processing target and acquiring language information therefrom;

25 outline estimation means for estimating an outline of an intonation for each assumed accent phrase configuring the text based on the language information acquired by the text analysis means;

shape element selection means for selecting an intonation pattern from a database accumulating intonation patterns of an actual speech based on the outline of the intonation, the outline having been estimated by the outline estimation means;

5 shape element connection means for connecting the intonation pattern for each assumed accent phrase to the other, the intonation pattern having been selected by the shape element selection means, and generating an intonation pattern of an entire body of the text; and

10 speech generation means for generating the speech based on the intonation pattern generated by the shape element connection means.

25. The program according to claim 24, wherein, when the  
15 assumed accent phrase applies to a predetermined phrase of a speech recorded in advance, the outline estimation means realized by the program acquires information concerning an intonation of a portion of the phrase of the recorded speech, the phrase corresponding to the assumed accent phrase, from  
20 another database storing information concerning intonations of the recorded speech.

26. A program for controlling a computer to perform a text-to-speech synthesis, the program allowing the computer to  
25 function as:

text analysis means for analyzing text, which is an object of a processing and acquires language information therefrom; synthesis unit selection means for selecting a waveform

element for a synthesis unit of the text; and

speech generation means for generating a synthesized speech by coupling the waveform element selected by the synthesis unit selection means to the other,

5 wherein the synthesis unit selection means selects the waveform element for the synthesis unit of the text, the synthesis unit corresponding to a boundary portion of a speech\_recorded in advance, from a first database which stores information concerning speech characteristics and a second database which  
10 stores information concerning a waveform of the speech recorded in advance.

27. A recording medium recording, to be readable by a computer, a program for controlling the computer to perform a  
15 text-to-speech synthesis,

wherein the program allows the computer to function as:  
text analysis means for analyzing text, which is an object of a processing and acquiring language information therefrom;  
outline estimation means for estimating an outline of an  
20 intonation for each assumed accent phrase configuring the text based on the language information acquired by the text analysis means;

shape element selection means for selecting an intonation pattern from a database accumulating intonation patterns of an  
25 actual speech based on the outline of the intonation, the outline having been estimated by the outline estimation means;

shape element connection means for connecting the intonation pattern for each assumed accent phrase to the other,

the intonation pattern having been selected by the shape element selection means, and generating an intonation pattern of an entire body of the text; and

5 speech generation means for generating the speech based on the intonation pattern generated by the shape element connection means.

28. The recording medium according to claim 27, wherein, when the assumed accent phrase applies to a predetermined phrase 10 of a speech recorded in advance, the outline estimation means realized by the program acquires information concerning an intonation of a portion of the phrase of the recorded speech, the phrase corresponding to the assumed accent phrase, from another database storing information concerning intonations of 15 the recorded speech.

29. A recording medium recording, to be readable by a computer, a program for controlling the computer to perform a text-to-speech synthesis,

20 wherein the program allows the computer to function as: text analysis means for analyzing text, which is an object of a processing and acquires language information therefrom; synthesis unit selection means for selecting a waveform element for a synthesis unit of the text; and

25 speech generation means for generating a synthesized speech by coupling the waveform element selected by the synthesis unit selection means to the other,

wherein the synthesis unit selection means selects the

waveform element for the synthesis unit of the text, the synthesis unit corresponding to a boundary portion of a speech recorded in advance, from a first database which stores information concerning speech characteristics and a second database which  
5 stores information concerning a waveform of the recorded speech.